

## Tilburg University

### Meaning and Intonation

Krahmer, E.J.; Swerts, M.G.J.

*Published in:*

Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4), january 19-12-2001, Tilburg

*Publication date:*

2001

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Krahmer, E. J., & Swerts, M. G. J. (2001). Meaning and Intonation: The cases of contrastive intonation and meta-linguistic negation. In H. Bunt, I. van der Sluis, & E. Thijsse (Eds.), *Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4), january 19-12-2001, Tilburg* (pp. 202-226). Tilburg University.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Meaning and Intonation: The cases of contrastive accent and meta-linguistic negation

Emiel Krahmer\* and Marc Swerts\*,<sup>‡</sup>

\* IPO, Center for User-System Interaction,  
TU/e, Eindhoven University of Technology,  
The Netherlands,  
{e.j.krahmer/m.g.j.swerts}@tue.nl

<sup>‡</sup> CNTS, Center for Dutch Language and Speech,  
UIA, University of Antwerp,  
Belgium

## Abstract

In this article we outline a methodological approach to the study of meaning and intonation. This approach focusses both on what speakers can produce (using production experiments) *and* on what hearers can perceive (using perception experiments). We show that such an experimental paradigm may yield interesting results from a semantical point of view by discussing two specific cases: contrastive accent and meta-linguistic negation. Concerning contrastive accents, we argue against the existence of a separately identifiable accent with a contrastive interpretation, though a contrastive intonation contour does appear to exist. We argue that this contour triggers a presupposition, which may be resolved using van der Sandt’s anaphoric theory of presuppositions. We also present empirical evidence for the existence of a set of prosodic differences between meta-linguistic negations and descriptive negations, a distinction which is the subject of considerable debate in presupposition theory. Finally, we argue that prosody gives rise to soft constraints, and point out that an optimality theoretic framework may be suitable to model the relation between prosody and meaning. In the discussion we outline some problems and prospects for such an account.

**Keywords:** contrastive accents, meta-linguistic negation, dialogue games, production and perception experiments, optimality theory

## 1 Introduction

In describing the sound shape of a language, it is common practice to distinguish between a segmental and a suprasegmental (or prosodic) level. The former refers to the individual speech sounds, seen as the basic units into which a continuous stream of speech can be subdivided. The latter comprises vocal features such as speech melody, tempo, loudness, pause, that are no typical attributes of the single segments, but are characteristic of longer stretches of speech. There has been a lot of research on how these two levels of sound structure may affect the meaning of an utterance. At the segmental level, one can view the individual speech sounds

as the basic building blocks out of which meaningful units are constructed. Though they have no intrinsic meaning of their own, they may change meaning in a discrete way, as the replacement of one phoneme by another creates a completely new word (the distinctivity principle). Consider the following pair of utterances:

- (1) a. John owns dogs.  
b. John owns hogs.

The difference in meaning between these utterances is clear. There is a clear-cut segmental contrast between the phonemes /d/ and /h/, which implies a categorical difference between the words *dogs* and *hogs*, and thus account for the difference in (truth-conditional) meaning between (1.a) and (1.b). A linguistic description of such phonological contrasts is helped by the existence of a lexicon which provides a yardstick to decide whether or not a difference in form leads to a difference in meaning.

Similar attempts to relate form to meaning at the suprasegmental level have often been less successful, because prosodic variation is usually not distinctive in this structural linguistic sense. It is generally more difficult to paraphrase how the meaning of an utterance is affected by replacing its intonation contour by another. For instance, consider the following variants of (1.a). In (2.a), the word *dogs* is pronounced with a sharp rise in pitch (an H\* pitch accent in the terminology of Pierrehumbert 1980), while in (2.b) it is pronounced with a lower-rising pitch accent (notated as L+H\*).

- (2) a. John owns DOGS  
H\*  
b. John owns DOGS  
L+H\*

What is the difference in ‘meaning’ between (2.a) and (2.b)? For instance, what is the function of the L+H\* accent in (2.b)? The literature contains at least the following, partially overlapping suggestions. According to Pierrehumbert & Hirschberg (1990) it marks a contrastive relation between *dogs* and something else. Vallduví (1990) claims that it indicates that the NP *dogs* is a link (an instruction to update a file card). According to the theory of Hendriks and Dekker (1995) it means that *dogs* is a non-monotone anaphor, while Steedman (2000) would claim that it indicates theme-hood. To make things even worse, it is still a matter of considerable debate what the descriptive intonational units are (for instance, whether a separately identifiable L+H\*-form accent does exist, see e.g., Steedman 2000), whether the assumed meaning of a contour generalizes to all tokens of that intonation pattern, how one should account for the variability between speakers in how they supplement their utterances with intonation patterns and for the variability between listeners in how they interpret particular contours, and how one should deal with the fact that the linguistic and situational context of an utterance may overrule the meaning of a given intonational contour.<sup>1</sup> The key problem seems to be that prosody often involves gradient rather than categorical differences, which is of course a severe complication when one wants to apply the principle of distinctivity to prosodic features, and assign semantic properties to these features.

---

<sup>1</sup>Moreover, it is worth stressing that prosody may also be ‘meaningful’ in quite different ways, to signal communicatively relevant phenomena like the cocktail party phenomenon, turn-taking, emotional and attitudinal aspects of utterances, etc.

In our opinion, the best methodological way to resolve such uncertainties regarding the interrelations between meaning and intonation is by doing experiments. In this article we outline such an experimental approach, focussing both on what speakers can produce (using production experiments) *and* on what hearers can perceive (using perception experiments). These perception experiments are particularly interesting because they explicitly trade on the assumption that meaning distinctions are only communicatively relevant if they can reliably and consistently be ‘interpreted’. We illustrate the experimental approach for two notorious cases; namely contrastive accents and meta-linguistic negation. Regarding contrastiveness, there is a general consensus in the literature that contrastive meaning can be signaled by means of accent distribution. For instance, what is conveyed by the following example is that the speaker bought the poodle which is white, and not any other-colored poodle which the addressee might have in mind.

(3) I bought the WHITE poodle.

However, the existence of *additional* phonological features which distinguish contrastive accents from more ‘neutral’ accents only marking new information is hotly debated. Some maintain that contrastive accents *are* formally different from other accents, either because the *type of accent* is different for the contrastive cases or because they are more *prominent*. Some people mention the existence of a sudden drop in pitch after the contrastive accent, whereas a non-contrastive accent is more likely to be sustained. Pierrehumbert & Hirschberg (1990) suggested that contrastive accents have an L+H\* pattern while novelty accents have an H\* form. Bartels & Kingston (1994) were unable to find support for Pierrehumbert & Hirschberg’s suggestion, but found evidence instead that contrastive accents tend to have higher peak heights than novelty accents (cf. also Ladd 1983). Others, however, maintain that contrastive accents do *not* exhibit specific intonation features. This is the position taken by intonologists like Halliday and Bolinger. “*As far as we can tell from the behaviour of pitch, nothing is uniquely contrastive*” (Bolinger 1986:342). The experimental evidence presented here will suggest that a separate contrastive accent does not exist (in Dutch), though a contrastive *intonation contour* does. We propose to interpret this contour as the trigger of a presupposition, to be resolved in the manner of van der Sandt (1992).

The difference between ‘ordinary’ (or ‘descriptive’) negation and ‘meta-linguistic’ (or ‘corrective’) negation has been widely studied in the context of presupposition theory. Consider:

- (4) Q: Is the king of France bald?  
 A<sub>1</sub>: No, because there is no king of France.  
 A<sub>2</sub>: No, he isn’t.

Here the two disconfirmations (A<sub>1</sub> and A<sub>2</sub>) perform rather different functions: the meta-linguistic negation in A<sub>1</sub> explicitly denies the underlying presupposition of Q’s question, while the ordinary negation A<sub>2</sub> does not, but only points out that the king of France still has his own hair. In partial approaches to presupposition (see e.g., Beaver and Krahmer 2000 for an overview), this problem is often solved by defining two distinct negation operators and letting natural language negation be ambiguous between the two.<sup>2</sup> This often meets with the criticism that there is no linguistic evidence for such an ambiguity (see e.g., Horn 1985).

---

<sup>2</sup>In fact, Beaver and Krahmer (2000) present an alternative to postulating undesired ambiguities, which uses Bochvar’s (1939) assertion operator *A* as a *presupposition wipe-out device*, in the sense that whatever is presupposed by a partial logical formula  $\varphi$ ,  $A\varphi$  presupposes nothing.



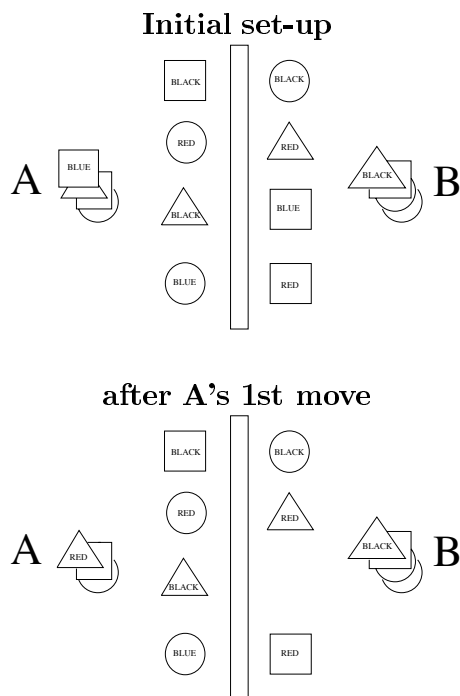


Figure 1: Bird's-eye view of the experimental set up. The initial stage and the first move of the dialogue game are displayed.

It has been suggested that there might be intonative differences between the two types of negation (e.g., by Beaver 1997). We shall present empirical evidence which suggests that this is indeed the case.

This paper is explicitly aimed at making our experiments and results accessible for semanticists. For the gory statistical details and the (psycho-)acoustical nitty gritty we refer to Krahmer and Swerts (2000) and Krahmer et al. (2000). In this article we focus on the semantic ramifications of the experimental results presented there, and on the relation between meaning and intonation in general.

## 2 Contrastive accents

In this section the alleged existence of contrastive accents is studied, both from a speaker's perspective (section 2.1) and a hearer's perspective (section 2.2), followed by a general discussion (2.3) in which a semantic analysis of contrastive intonation is proposed.

### 2.1 Speaker's Perspective: Production Experiment

**Method** The production experiment, which tries to mimic natural conversation, consists of a simple dialogue game played by four pairs of subjects. The game is essentially an alignment task of figures played by two subjects, call them A and B, who are separated from each other by a screen. Figure 1 gives a bird's-eye view on the experimental set-up of the elicitation task, both the initial stage and the stage after A's first move. In each game, both players

Table 1: Examples of the four contexts

NN	(beginning of game) B: “blue square”
CC	A: “red circle” B: “blue square”
CG	A: “yellow square” B: “blue square”
GC	A: “blue triangle” B: “blue square”

had an identical set of eight cards to their disposal, every card showing a geometrical figure (square, triangle or circle) in a particular color (blue, red, black or yellow). Beforehand, the eight cards were divided in two sets of four cards each: one set of four cards was ordered as a stack, the other half was unordered. The stacks of A and B were disjoint: thus A’s stack corresponded with the unordered set of B, and *vice versa*. At the onset of the game, one participant, say A, describes the figure on the top of his stack (“a blue square”) and this prompts both A and B to place the card with the blue square at the first position of the row of figures under construction. Next, B takes over and describes the object on top of *his* stack (“a black triangle”). Now, both A and B place the card with this object on the second place in the row with figures, and so on. The game is over when both players are out of cards. There are no winners or losers.

The data thus obtained allow an unambiguous operationalization of the relevant contexts. A property (colour or figure) is defined to be *new* (N) to the conversation if it is mentioned in the first turn of the current dialogue game, it is *given* (G) if it was mentioned in the previous turn and finally a property is *contrastive* (C) if the object described in the previous turn had a different value for the relevant property. By systematically varying the sequential order of the cards in front of the subjects, target descriptions were collected for the eight speakers in four contexts: no contrast (all new, NN), contrast in the adjective (CG), contrast in the noun (GC), all contrast (CC). Table 1 summarizes the situation.

**Results** Concerning the speaker perspective two questions are relevant: (1) which words received an accent in which contexts and (2) are these accents of a different type. Ad (1): All utterances of two target descriptions (“blue square” and “red square”) were used for a distributional analysis performed by two intonation experts. Table 2 summarizes the results and reveals a clear trend: in the NN (no contrast/all new case) both adjective and noun are (nearly) always accented, and in most cases the same holds for the CC (double contrast) cases. When one item is given, while the other is contrasted (i.e., the CG and GC cases), the contrasted item generally is the only accented word. Even though both CG & GC, and NN & CC are strikingly similar, there are two exceptions. First, there is a complete lack of postnuclear accents in the CG case, while occasionally prenuclear accents on the adjective occur in the GC case.<sup>3</sup> Second, CC differs from NN in that there are a number of utterances

---

<sup>3</sup>With the ‘nuclear accent’ we refer to the most prominent accent in an intonative phrase (for English and

Table 2: Accent distribution according to two intonation experts (exp<sub>1</sub> and exp<sub>2</sub>) on all target utterances “blauw vierkant” and “rood vierkant” (blue and red square respectively) in four contexts: NN (no contrast), CC (all contrast), CG (contrast only in adjective), GC (contrast only in noun). One CG utterance is missing.

Context	Accent on					
	Adj Only		Noun Only		Adj and Noun	
	exp <sub>1</sub>	exp <sub>2</sub>	exp <sub>1</sub>	exp <sub>2</sub>	exp <sub>1</sub>	exp <sub>2</sub>
NN	0	0	0	2	16	14
CC	3	3	2	4	11	9
CG	15	15	0	0	0	0
GC	1	1	11	11	4	4

in the CC context with an accent only on the adjective or the noun. Looking at these exceptional cases revealed that in all cases the speaker made a contrast with his or her *own* last utterance, thereby ignoring their partners last contribution. Interestingly, all these “ego-centric” speakers happen to end their utterances on a high (H%) boundary tone, whereas the other speakers uniformly employed low (L%) boundary tones. Apparently, in the current experiment, speakers using high boundary tones signal that they want to continue their own train of thought, which leads to what is generally referred to as list intonation.

Now to question (2), do the contrastive accents have a specific intonational shape. If one makes the common assumption that a single accent on the noun is ambiguous between a broad focus and a narrow focus reading, then one might expect that a contrastive accent manifests itself most clearly in the noun position. However, for none of our speakers does a comparison of a single contrastive accent on the noun (GC) with a newness accent on that same syntactic position reveal differences with respect to the type of accent. Interestingly, at first sight the single contrastive accent on the adjective (CG) *is* of a different type than the newness accent on that same syntactic position. However, the single contrastive accent on the adjective is of the *same* type as the accent on the noun. Thus: the difference in type of accent is only apparent, since in the CG context the adjective is associated with a nuclear accent in a non-default position.<sup>4</sup> So, as far as *type* of accent is concerned there seems to be no difference between contrastive and newness accents. This does not necessarily mean that hearers are not able to distinguish the two. It might be, for instance, that contrastive accents stand out perceptually. To find out, a perception experiment is called for.

## 2.2 Hearer’s Perspective: Perception Experiment

**Method** In the literature it has been suggested that contrastive accents stand out, because they are more prominent (extra high or boosted, see e.g., Ladd 1983, Bartels & Kingston 1994). The goal of this perception experiment is to find out whether hearers are indeed able to distinguish contrastive accents from other accents. Perceptual data were obtained from sixteen prosodically naive subjects (distinct from the eight speakers) who participated in an individually performed listening task. From the eight speakers in the production experiment

---

Dutch this is often claimed to be the *last* major accent in a phrase).

<sup>4</sup>We refer to Krahmer and Swerts (2000) for a more detailed analysis and the associated sound files.

Table 3: Results of the perceptual measurements of prominence ratings on Adj and Noun for speaker JR in the isolated and complete conditions. The four contexts CG, NN, CC and GC are defined in the text, for readability the word of interest is capitalized.

	Adj				Noun			
	Gc	Nn	Cc	Cg	cG	nN	cC	gC
<i>complete</i>	1	57	44	90	10	52	44	86
<i>isolated</i>	6	86	26	74	14	89	19	70

two were selected: JR (a low-ending speaker) and WY (a high ending speaker). It is interesting to take these two because they use essentially different intonation contours, which might effect the perceived prominence of the accents (the former employs a rising contour, the latter not).

The subjects of the perception experiment were presented with four lists of 12 pairs of phrases: two lists of utterances produced by speaker JR, and two lists by speaker WY. The data were presented in two conditions: *complete* (entire utterances) and *isolated* (words). The *rationale* for these two conditions is the following: if a contrastive accent really stands out (i.e., is perceptually distinguishable from a more neutral, newness accent), then this should be a property of the accent itself, and thus this should hold *both* for the complete condition *and* for the isolated one. In the complete condition, subjects could hear the utterances as they were originally produced by speaker JR or WY. In the isolated condition, listeners were presented with one word cut from its context, either the adjective or the noun. The listener was instructed (for each condition) to select the member of the pair which he or she thought was the most prominent: in the complete condition, they were asked to focus on either the noun or the adjective and to determine by forced choice which of the pairs contained the most prominent one. In the isolated condition, they had to select (again by forced choice) the word which they judged to be the most prominent. No specific definition of prominence was given to the subjects. Given that the utterances were elicited in four contexts (NN, CC, GC, CG), subjects had to make  $3! = 6$  pairwise comparisons (NN-CC, NN-GC, NN-CG, CC-GC, CC-CG and GC-CG). These 6 pairs were always presented in both orders (A-B and B-A), which leads to 12 pairwise comparisons per list. The listener heard each of these 12 pairs two times in a row, before he or she had to make a judgment on the pair. Both the order of the pairs in a list as the order of the different lists were fully randomized.

**Results** Tables 3 and 4 give an overview of the perception results for speakers JR and WY respectively. The tables sum over the number of times a word (either the adjective or the noun) was judged the most prominent in a pairwise experiment. The maximum here is 96 ( $16 \text{ subjects} \times 2 \text{ judgments} \times 3 \text{ comparisons}$ ). Note that these figures only give the overall results, and that the details of the pairwise comparisons are lost in this presentation (but see Krahmer and Swerts 2000). In the following, capitalized letters indicate the words on which the subject scored (thus: ‘Gc’ indicated that the subject had to rate the prominence of the adjective in GC context; the adjective is given while the noun is contrastive). The results thus obtained for the complete condition are basically the same in all cases. That is: in all four cases, a single contrastive accent (gC or Cg) is judged to be the most prominent, while givenness (Gc or cG) uniformly scores lowest on prominence. The double contrast (cC/Cc)

Table 4: Results of the perceptual measurements of prominence ratings on Adj and Noun for speaker WY in the isolated and complete conditions. The four contexts CG, NN, CC and GC are defined in the text, for readability the word of interest is capitalized.

	Adj				Noun			
	Gc	Nn	Cc	Cg	cG	nN	cC	gC
<i>complete</i>	8	33	64	87	4	47	60	81
<i>isolated</i>	0	39	77	76	16	27	86	63

and no contrast/all new (nN/Nn) cases are in between these two extremes, with for WY double contrast being judged relatively more prominent than the no contrast case, whereas the trend is in the opposite direction for JR. Thus both adjective and noun lead to comparable prominence patterns. This is true for both speakers, even though they provided their utterances with different intonation contours (low vs. high ending). A  $\chi^2$  test revealed that all the different distributions for the data obtained in the complete condition are significantly different from chance ( $p < 0.001$ ).

As said, if contrastive accents as such stand out perceptually with respect to newness accents, than we expect this to be the case in both the complete and the isolated condition. However, when focussing on the isolated condition the overall picture changes dramatically. Within speakers essentially the same pattern for the adjective and the noun can be observed, but this pattern is rather different from the complete pattern. In particular, for JR the newness accent (Nn/nN) is suddenly judged to be the most prominent, while the double contrast accent (Cc/cC) scores almost as low as the given case (Gc/cG). For WY, the single contrast (Cg/gC) gets comparatively lower prominence ratings. The distributions in the isolated condition are again significantly different from chance level, according to the  $\chi^2$  test (again  $p < 0.001$ ). Also, for each speaker, the distributions for the isolated condition are significantly different from those obtained in the complete condition (the Adj for WY:  $p < 0.025$ , the remaining three cases  $p < 0.001$ ).

The results are clear: in the complete condition, single contrastive accents stand out as the most prominent ones, irrespective of the intonation contour (high vs. low ending) and irrespective of the place of the accent in the utterance (adjective vs. noun). Similarly, given items are always judged to be the least prominent, while the all new and double contrast cases lie in between the two extremes. It is striking that in the isolated condition a different picture emerges, in that, for instance, the single contrastive accents are no longer perceived as being the most prominent ones. The only difference between the isolated and the complete condition is that in the former condition words are taken from their natural context. This suggests that prosodic context information plays a central role in the complete condition, whereas in the isolated condition hearers solely base prominence judgments on acoustic properties of the target word (in particular pitch and loudness).

### 2.3 Discussion

Summarizing the main results of section 2: no evidence is found that “contrastive” accents have a distinctive *shape*, which sets them apart from other accents. The apparent difference between a “new” and a “contrastive” accent on the adjective is actually a difference between

a prenuclear and a nuclear accent; because of deaccentuation<sup>5</sup> of the noun, the “contrastive” accent on the adjective has become the final accent in a phrase, and thus gets a nuclear shape. In other words: apparently, the contrastive interpretation is not associated with a specific prosodic shape but rather with the non-default position of the nuclear accent. Concerning the issue of *prominence* differences, it appears that contrastive accents are perceived as more prominent than newness accents on the same syntactic position. However, this only holds true if subjects can listen to the *complete* utterance. The difference in perceived prominence tends to disappear if the noun or adjective is presented in isolation. This effect might be called *prosodic masking*: an isolated pitch peak is perceived as more prominent than the same peak presented in the context of an intonationally comparable pitch peak. (The Mt. Everest would be perceived as higher when encountered in the low lands than in the Himalaya.) For the prominence judgments, it seems that the prosodic context (whether or not the relevant accent is preceded or followed by another accent) is the major factor contributing to the perception of a contrastive intonation.<sup>6</sup>

What is the impact of the current results for semantics? It has been argued (e.g., by Cutler 1977) that context largely determines the interpretation of a given intonational pattern.<sup>7</sup> Interestingly, recent years have seen an increasing awareness of the influence of context on meaning, and this has prompted various semanticists to propose formal, context-dependent interpretations associated with contrastive accents (e.g., Rooth 1992, Hendriks and Dekker 1995, van Deemter 1999, Piwek 1998). Here, a different, more fundamental question was addressed, namely whether a separately identifiable contrastive intonation exists in the first place. Even though no acoustic evidence for a separate contrastive accent was found, the data show that contrastiveness *can* be determined on the basis of intonation. This means that we are now in a position where it makes sense to address the question what the meaning of contrastive intonation is. On a purely pre-theoretical level, it seems reasonable to assume that speakers signal a contrast relation to enhance the hearer’s processing of their utterance: by marking the information which is contrastive, e.g., “BLUE square,” they seem to say: pick the ‘referent’ of the previous square which we discussed and —for the current referent— modify the colour value by setting it to blue (compare Pechmann 1984).

Arguably, the presuppositions as anaphors theory of van der Sandt (1992) is well-suited to model this intuitive idea. We propose to associate a phrase like “a BLUE square” with a presupposition stating that there exists a differently coloured square. According to the theory of van der Sandt (1992) this presupposition needs to be resolved with respect to the preceding context (that is, the context should contain an ‘antecedent’ for the presupposition). This implies that the direct linguistic context, i.e., the previous turns, should contain a mention

---

<sup>5</sup>With deaccentuation we refer to the phenomenon that a word which normally would be accented does not receive a pitch accent in a particular context.

<sup>6</sup>Arguably, one limitation of the perception test described here is that it is concerned with ‘prominence’ and not with contrastiveness directly. However, in our opinion the latter issue presupposes the former. That is: before we can associate contrastive interpretations to accents, it should be established that contrastive accents can be distinguished (*as accents*) from other accents. The perception test described here illustrates that a contrastive intonation contour (but not a contrastive accent as such) appears to exist. We are currently working on a different perception experiment, where listeners are asked to reconstruct the preceding utterance on the basis of complete utterances. The results of a pilot test indicate that subjects are able to do this only in the case of contrastiveness (Swerts and Krahmer, in progress). Interestingly, it appears that subjects find this easier when the adjective is contrastive than when the noun is contrastive, which gives an interesting twist to the narrow focus/broad focus debate.

<sup>7</sup>Cutler (1977:106): “(...) *the attempt to extract from [intonation contours] an element of commonality valid in all contexts must be reckoned a futile endeavour.*”

of such a square. More precisely, if the utterance ends on a low boundary tone (L%) the antecedent is to be found in the previous turn, if it ends on a high boundary tone (H%) it is to be found in the turn before the previous one (i.e., the previous turn of the current speaker). If this is not the case, the prediction is that the utterance of “a BLUE square” was infelicitous. An account along the lines sketched here would also model the finding that neither the distributional nor the perceptual analysis revealed essential differences between the NN and the CC contexts: a double contrast would have very little informative content for the hearer as it would urge her to look for an antecedent which has a different shape and a different colour, which is nearly tantamount to creating an entirely new object and as such has very little descriptive content.

### 3 Meta-linguistic negation

Let us now turn to meta-linguistic negation. Once again, we shall study this phenomenon from both a speaker perspective (3.1) and a hearer perspective (3.2), followed by a general discussion in section 3.3.

#### 3.1 Speaker’s Perspective: Production Experiment

**Method** How do speakers produce meta-linguistic negations, and is there a difference with ‘ordinary’, descriptive negations? To address this question, we first need to come up with an independent criterion to distinguish between the two. According to Horn (1985:136), metalinguistic or cancelling negation expresses something like “I object to *u*”, where *u* is crucially a linguistic utterance rather than an abstract proposition”. Consider the following rule from Groenendijk et al. (1996):

**Rule H2** If a sentence is uttered which is incompatible with a participant’s information state, then she does not update with it, but signals the incompatibility by uttering a sentence that contradicts the sentence uttered.

We take it that meta-linguistic negations are examples of ‘sentences contradicting the previously uttered sentence’. They function as a negative, ‘go back’ signal, indicating that there is an apparent communication problem; a discrepancy between the last utterance of the addressee and the information state of the current speaker. If there are no communication problems, the speaker sends a positive, ‘go on’ signal. Our hypothesis is that speakers use more prosodically marked features in the case of ‘go back’ signals (indicating a communication problem) than in the case of ‘go on’ signals.

To address this question a corpus of human-machine dialogues was used.<sup>8</sup> This corpus consists of 120 dialogues with two speaker-independent Dutch spoken dialogue systems which provide train time table information (see Weegels 1999). The systems prompt for unknown slots, such as departure station, arrival station, date, etc., in a series of questions. Twenty subjects were asked to query both systems via telephone on a number of train journeys. They were asked to perform three simple travel queries on each system (in total six tasks). In the corpus used in this study, subjects may use disconfirmations in response to two kinds of questions, of which (5.a) and (5.b) are representative examples.

---

<sup>8</sup>The current discussion of meta-linguistic negation is part of a wider research programme to study communication problems in human-machine conversation.

- (5) a. So you want to go from Eindhoven to Swalmen?  
 b. Do you want me to repeat the connection?

Both (5.a) and (5.b) are yes/no questions and to both “no” is a perfectly natural answer. However, the two questions serve a rather different goal, and consequently the corresponding negations have a rather different function. Question (5.a) is an (explicit) attempt of the system to verify whether its current assumptions (about the departure and arrival station) are compatible with the intentions of the subject.<sup>9</sup> If this is not the case, the subject will signal this (in line with rule H2 above) using a meta-linguistic negation, thereby indicating that at least one of the system’s assumptions is incorrect (*No, I do not want to go to Swalmen; I want to go to Reuver*).<sup>10</sup> Question (5.b), on the other hand, is not an attempt of the system to verify its assumptions, and hence it cannot represent incorrect system assumptions. A subsequent “no” answer from a subject thus serves as an ‘ordinary’ negation. So, the two kinds of system yes/no questions allow for an unambiguous distinction between normal and meta-linguistic negation. The respective disconfirmations, being lexically similar but functionally different, constitute minimal pairs, allowing us to check whether the various occurrences of this kind of utterance vary prosodically as a function of their context. In this way, they form ideal, naturally occurring speech materials for investigating the role of prosody.

For studying the speaker’s perspective a random selection of 109 negative answers to yes/no questions was taken from the corpus. If a negative answer follows a verification question (such as (5.a)), the subject’s utterance indicates that there are communication PROBLEMS. If a negative answer follows a standard yes/no question (like (5.b)) there are no communication problems (notated as  $\neg$  PROBLEMS).

Regarding their structure, the subjects’ negations were divided into three categories: (1) responses consisting of an explicit disconfirmation marker “no” (“nee”) only (we shall refer to these cases as ‘single no’), (2) responses consisting of an explicit disconfirmation marker followed by other words (‘no+stuff’ in the terminology of Hockey et al. 1997), (3) responses containing no explicit disconfirmation marker (‘stuff’).<sup>11</sup> The proportion of meta-linguistic negations is .62. The subjects’ responses to the yes/no questions were analysed in terms of the following features: (1) type of boundary tone in “no” (high or not high); (2) duration (in ms) of “no”; (3) duration (in ms) of pause after “no” before stuff; (4) duration (in ms) of pause between system’s prompt and user response; (5)  $F_0$  max<sup>12</sup> (in Hz) at energy peak of major pitch accent in stuff; (6) number of words in stuff.

**Results** Table 5 gives the distribution of different types of negations following either an unproblematic system utterance or one which contains one or more problems. A  $\chi^2$  test reveals that these numbers significantly differ from chance level ( $p < 0.001$ ). First, this table

<sup>9</sup>Due to the imperfectness of automatic speech recognition technology, current state dialogue systems are in constant need of verification.

<sup>10</sup>The example discussed in the introduction (4) is special in the sense that the meta-linguistic negation expresses an objection to a *presupposition* (namely that there is a king of France). Such examples with failing presuppositions have been the subject of long and difficult debates, and intuitions reported in the literature strongly covary with the theories of presupposition (Russellian or Strawsonian) one adheres to. Disconfirmations in reply to examples such as (5.a) are meta-linguistic in the sense of Horn (1985), but are not subject to the difficulties of failing presuppositions.

<sup>11</sup>As we shall see, meta-linguistic negations may occur which do not contain an explicit negation. An example would be: *A: Thomas ate some cookies. B: He ate ALL cookies!*.

<sup>12</sup> $F_0$  stands for *fundamental frequency*; changes in the fundamental frequency are the most commonly used approximation of perceived pitch variations.



Table 5: Numbers of negative answers following an unproblematic system utterance ( $\neg$  PROBLEMS) and following those containing one or more problems (PROBLEMS).

Type	$\neg$ PROBLEMS	PROBLEMS	TOTAL
no	18	11	29
stuff	0	24	24
no+stuff	23	33	56
TOTAL	41	68	109

Table 6: Presence or absence of high boundary tones following occurrences of “no” (single no and no+stuff) for positive and negative cues.

High boundary tone	$\neg$ PROBLEMS	PROBLEMS	TOTAL
Absent	32	7	39
Present	9	37	46
TOTAL	41	44	85

shows that the minimal response, a single no, is in the majority of the cases used when there are no communication problems. Second, single stuff responses are exclusively reserved for responses following a system utterance with one or more problems. The majority of the responses to yes/no questions in our data, however, is of the no+stuff type, which may serve either as a descriptive or as a meta-linguistic negation. The lexical material in the stuff is quite different for the two signals: for the positive cases, the subsequent words are mostly some polite phrases (“thank you”, “that’s right”); for the meta-linguistic cases, the stuff usually is an attempt to correct the information which caused the problems.

Table 6 displays the presence or absence of high boundary tones (H% in the terminology of Pierrehumbert 1980) on the word “no” (for the single no and no+stuff cases). A  $\chi^2$  test reveals that this distribution is again well above chance level ( $p < 0.001$ ). In the case of problems, the “no” is generally provided with a question-like H% boundary tone, which is absent when “no” follows an unproblematic system utterance.

The results for the continuous prosodic features of interest are given in table 7. Taking the utterances of all subjects together, a t-test reveals a significant difference for each of these features. The trend is the same in all cases: corrective, meta-linguistic negations are comparatively marked. First, the word “no” —when it occurs— is longer. Second, there is a longer delay after a problematic system prompt before subjects respond. Third, in the no+stuff utterances, the interval between “no” and the remainder of the utterance is longer. Fourth, the stuff part of the answer usually contains a high-pitched accent to mark corrected information, whereas in the unproblematic case the stuff is usually prosodically unmarked. Finally, the stuff part tends to be longer in number of words. In sum: there are clear prosodic differences between meta-linguistic and descriptive negations.

Table 7: Average values for various features. Duration of “no” (for all occurrences of “no”: single no and no+stuff), delay between end of system utterance and beginning of user’s disconfirmation (all cases), pause between “no” and stuff (for no+stuff cases),  $F_0$  max in stuff and number of words in stuff (both for no+stuff and stuff).

Feature	$\neg$ PROBLEMS	PROBLEMS
Duration of “no” (ms)**	226	343
Preceding delay (ms)**	516	953
Following pause (ms)*	94	311
$F_0$ max in stuff (Hz)*	175	216
Words in stuff**	2.61	5.42

\*\* $p < 0.001$ , \* $p < 0.05$

### 3.2 Hearer’s Perspective: Perception Experiment

**Method** It seems a reasonable hypothesis that when speakers systematically dress up their utterances with certain features, hearers will be able to attach communicative relevance to the presence or absence of these features. To test if this is indeed the case for the acoustic properties of utterances of “no” found in the study of the speaker’s perspective, a perception experiment was carried out. For this experiment we used 40 “no”s, all taken from no+stuff disconfirmations. We opted for no+stuff disconfirmations since these are the most frequent and are equally likely to be used following problematic and unproblematic system utterances from a distributional perspective (see table 5), and are thus least biased in terms of their function as positive or negative cues. For the perception study, we only used the “no”-part of these utterances, given that the stuff-part would be too informative about their function (compare answers A<sub>1</sub> and A<sub>2</sub> in (4)). Of the 40 “no”s, 20 functioned as a normal negation and 20 as a meta-linguistic negation. Subjects of the perception experiment were 25 native speakers of Dutch. They were presented with the 40 stimuli, each time in a different random order to compensate for any potential learning effects. They heard each stimulus only once. The experiment was self-paced and no feedback was given on previous choices. In an individual, forced choice task, the subjects were instructed to judge for each “no” they heard whether the speaker signaled a communication problem or not. They were not given any hints as to what cues they should focus on. The subjects were first presented with four “exercise” stimuli to make them aware of the experimental platform and the type of stimuli. It is worth stressing that the choice to use only “no”s extracted from no+stuff answers implies that not all the acoustic features which speakers employ (see above) survive in the current perceptual analysis. In particular, we lose the features delay (time between end of prompt and start of subject’s answer), pause (time between end of “no” and beginning of stuff) as well as potential cues in the stuff part (e.g., number of words, narrow-focused pitch accents).

**Results** Table 8 summarized the results of the perception experiment. Of the non-problem signalling negations 17 out of the 20 cases were classified by a significant number of subjects as cases in which the speaker did not signal a problem. The remaining three cases were in the expected direction, though not significant. Of the corrective negations, 15 out of 20 cases were classified correctly as instances of “no” signaling problems. Interestingly one meta-linguistic

Table 8: Perceived classification of positive and negative signals.

	Perceived as $\neg$ PROBLEMS	No significant difference	Perceived as PROBLEMS	Total
$\neg$ PROBLEMS	17	3	0	20
PROBLEMS	1	4	15	20
Total	18	7	15	40

negation was significantly misclassified as a descriptive signal. A post-hoc acoustic analysis of this “no” revealed that it shared its primary characteristics with ordinary descriptive negations. In particular: the “no” was relatively short, and lacked a high boundary tone. Table 8 clearly shows that subjects are good at correctly classifying instances of “no”, extracted from no+stuff utterances, as descriptive or meta-linguistic negations.

### 3.3 Discussion

Thus, both the production and the perception experiment strongly suggest that meta-linguistic negations are prosodically different from descriptive negations. In our opinion this is an important finding from a semantic perspective. It gives “ambiguists” (those who maintain that natural language negation is semantically ambiguous, in Horn’s 1985 terminology) an empirical argument for postulating a lexical ambiguity for negation. In a similar vein, it gives “monoguids” (such as Horn himself) an independent criterion for deciding whether a negation is used descriptively or meta-linguistically. Beaver & Krahmer (2000) propose to treat cases of presupposition denial (as  $A_1$  in (4)) using Bochvar’s assertion operator. The nice thing about this latter proposal is that it is not only applicable to negations, but also to other meta-linguistic cases such as the following (due to Beaver 1997 and Horn 1985 respectively).

- (6) a. If Mary KNOWS that Bill is happy, then I’m a Dutchman — she merely believes it.  
b. You did WHAT with Sally and Billy?

It would be highly interesting to find out whether some of the intonative properties which distinguish meta-linguistic from descriptive negation can be also be found in other meta-linguistic phenomena, although we do not immediately see how one could test this.

## 4 General discussion

The relation between meaning and intonation is a highly complex one. We have argued that to investigate this relation an experimental approach is called for, in particular one in which both the speaker’s perspective and the hearer’s perspective are taken into account. One obvious methodological advantage of doing experiments with different speakers and listeners is that one gains insight into inter- and intra-subject agreement, both in terms of production and perception, and that it provides a handle on how to deal with the intrinsic variability between subjects regarding intonational matters. In addition, it is instrumental in determining what is essential (that which many subjects agree on) and what is peripheral (those features regarding

which there is little consensus). It is difficult to imagine how this distinction can be made on the basis of researchers’ intuitions alone.

The motivation for looking both at speakers and listeners is that it does justice to our belief that a feature can only be communicatively relevant if it is not only encoded in the speech signal by a speaker, but if it can also be interpreted by a listener. There is an interesting parallel with Optimality Theory (OT) here; OT syntacticians tend to focus on the speaker perspective, while OT semanticists (following Hendriks and de Hoop *to appear*) tend to focus on the hearer. Recently, there has been an increased interest in combining the two perspectives (see e.g., Beaver 2000 for such a plea). In fact, we believe that an OT-like framework is eminently suitable to model the intricate relationship between intonation and meaning. First of all, it is clear that whatever meaning intonational contours may have, they can easily be ‘overruled’ by features from other linguistic levels or by the situational context. This has, for instance, been illustrated by Geluykens (1987) who showed that the classification of intonation contours as statements or questions is highly influenced by the lexical content of the utterances on which they occur. He tested this perceptually using sentences with a declarative syntax, finding that high-ending contours are more likely to trigger an interrogative interpretation if they occur on question-prone utterances like *You feel ill* than on statement-prone utterances like *I feel ill*.<sup>13</sup> Or take a sentence like “you fucking idiot” spoken to the driver of a car that just hit the speaker’s car. Whatever contour the speaker would put on that utterance, it will be difficult to seriously affect its intended basic meaning. In this article (section 3), we have seen that user responses to communication problems contain prosodic but also non-prosodic cues, like the lexical materials in the stuff part of the no+stuff utterances. These cues may even conflict, such as the particular example of the single meta-linguistic negation consistently classified as a descriptive negation (see table 8); even though the prosodic features of “no” suggested that there were no communication problems, this is overruled by the lexical material in the stuff (“not to Amsterdam, to Opdam!”).<sup>14</sup>

Note that for an OT approach which has something to say about the relation between intonation and meaning, it is essential to integrate different levels of linguistic analysis into a single tableaux. A plea for such an integrated approach can also be found, albeit for different reasons, in Beaver (2000), who states: *There are syntactic constraints, semantic constraints and relational constraints, but interesting OT theories of language will generally not be easily labeled as OT syntax or OT semantics. One thing that OT offers us is a new way of looking at the syntax-semantics-pragmatics interface which emphasizes the significance of relational constraints, and treats purely syntactic, purely semantic and purely pragmatic constraints as parochial special cases.* Following up on this, another thing that OT offers is a handle on intonational variation accross languages and its contribution to meaning. The idea is that many of the constraints are universal, although the ordering may differ accross languages. We have

---

<sup>13</sup>This difference trades on the assumption that is more easy to make statements about one’s own internal state than about those of other people.

<sup>14</sup>An open research question related to this would be how *non-categorical* features can be integrated in an OT approach. In principle OT constraints are universal restrictions (“all feet are right-aligned with the right edge of the word”, McCarthy & Prince 1993). However, as shown here, some aspects are of a scalar nature, such as gradient difference in pitch range (see section 2), and word duration and pauses (see section 3). Moreover, the more features apply, the stronger the semantic effect. Our view is that it may be possible to connect scalar values to differences in constraints (see also Boersma, 1998). For instance, prominence judgments seem to be gradient in that some differences in pitch range and loudness represent differences in cue strength, which influences the perception of accents, though it is unclear whether the gradient information is truly continuous or whether those types of continua can be divided into reliable categories which are invariant across speakers.

applied the same experimental methodology we used for analysing accent patterns in Dutch to Italian and Japanese (see Swerts et al. 1999, 2000). These studies clarify that the relative importance of intonation for signalling information status can differ considerably between languages. It appears that a language like Italian which is comparatively less flexible than Dutch in terms of modifying intonation patterns to signal information status may compensate for this by a more flexible word order and by subtle variations in pitch range. In Japanese, a language where pitch accent is lexically determined, there exist particular morphemes that can signal information status (such as *wa*), which takes over part of the intonative marking of information status in Japanese. The finding for the cases of meta-linguistic negation are different, in that the usage of marked prosodic features to signal communication problems appears to have a universal flavour. We can report similar effects in quite different types of human-human and human-machine interactions, collected for Japanese (Shimojima et al., 2000) and American English (Swerts et al., 2000). The fact that we find similar phenomena in different languages suggests the working of a universal constraint, which should be captured in an OT approach to intonation.

## References

- Bartels, C., Kingston, J., (1994). Salient pitch cues in the perception of contrastive focus. In: Bosch, P., van der Sandt, R. (Eds.), *Focus & Natural Language Processing*, IBM Working Papers, TR-80.94-006.
- Beaver, D. (1997), Presupposition. In: *Handbook of Logic and Language*, J. van Benthem & A. ter Meulen, Elsevier Science Publishers, 939-1008
- Beaver, D. and E. Krahmer, (2000), A partial account of presupposition projection *Journal of Logic, Language and Information*, in press.
- Beaver, D. (2000), *Centering and the Optimization of Discourse*, manuscript, Stanford University.
- Bochvar, D., (1939), Ob odnom trehznachom iscisenii i ego primeneii k analizu paradoksov klassickogo rassirennogo funkcional ‘nogo iscisenija’. In: *Matematiciskij sbornik*, 4 (English translation (1981): “On a Three-valued Logical Calculus and Its Applications to the Analysis of the Paradoxes of the Classical Extended Functional Calculus.” In: *History and Philosophy of Logic* 2:87-112)
- Boersma, P., (1998), *Functional Phonology*, PhD. Dissertation, University of Amsterdam.
- Bolinger, D., (1986). Intonation and its Parts. London, Edward Arnold.
- Cutler, A., 1977. The context-dependence of “intonational meanings”. In: Papers from the 13th regional meeting. Chicago, Chicago Linguistic Society, pp. 104-115.
- van Deemter, K., 1999. Contrastive stress, contrariety and focus, *Focus: Linguistic, Cognitive and Computational Perspectives*, Bosch, P., van der Sandt, R. (eds.), Cambridge University Press.
- Geluykens, R., (1987), Intonation and speech act type, *Journal of Pragmatics*, 11: 483-94.
- Groenendijk, J., M. Stokhof, F. Veltman, (1996), Coreference and modality in the context of multi-speaker discourse. In: *Context dependence in the analysis of linguistic meaning*. H. Kamp and B. Partee (eds.), Stuttgart, IMS, 195-216.
- ’t Hart, H., R. Collier and A. Cohen (1990), *A Perceptual Study of Intonation: An experimental-phonetic approach to speech melody*, Cambridge University Press.
- Hendriks, H. and P. Dekker (1995). Links without locations. In: Dekker, P., Stokhof, M. (Eds.), *Proceedings of the Tenth Amsterdam Colloquium*, Amsterdam, ILLC, pp. 339-350.

- Hendriks, P. and H. de Hoop, to appear. Optimality Theoretic Semantics, *Linguistics and Philosophy*.
- Hockey, B., D. Rossen-Knill, B. Spejewski, M. Stone and S. Isard, (1997), Can you predict answers to y/n questions? Yes, no and stuff. In: *Proceedings Eurospeech'97*, Rhodes, 2267-2270.
- Horn, L. (1985), Meta-linguistic negation and pragmatic ambiguity, *Language*, **61**:121-174.
- Krahmer, E. and M. Swerts (2000), On the alleged existence of contrastive accents, *Speech Communication*, in press.
- Krahmer, E., M. Swerts, M. Theune and M. Weegels (2000), The Dual of Denial: Two uses of disconfirmations in dialogue and their prosodic correlates, *Speech Communication*, to appear.
- Ladd, D., (1983). Phonological features of intonational meaning. *Language* **59**: 721-759.
- McCarthy, J. & A. Prince (1993), Generalized Alignment. in: G. Booy and J. van Marle (eds.), *Yearbook of Morphology*, 1993, Dordrecht, Kluwer Academic Publishers, 79-153.
- Pechmann, T., (1984) Überspezifizierung und Betonung in Referentieller Kommunikation. Ph.D. dissertation, Mannheim University.
- Piwek, P., 1998. *Logic, Information & Conversation*, Ph.D dissertation, Eindhoven.
- Pierrehumbert, J. (1980), *The Phonology and Phonetics of English Intonation*, Ph.D Dissertation, MIT.
- Pierrehumbert, J. and J. Hirschberg (1990), The meaning of intonational contours in the interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M. (Eds.), *Intentions in Communication*, Cambridge MA: MIT Press, pp. 342-365.
- Rooth, M., 1992. A theory of focus interpretation, *Natural Language Semantics* **1**: 75-116.
- van der Sandt, R., (1992), Presupposition Projection as Anaphora Resolution, *Journal of Semantics* **9**: 333-377
- Shimojima, A., Y. Katagiri, H. Koiso and M. Swerts (2000), The informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. *Speech Communication*, to appear.
- Steedman, M. (2000), Information Structure and the Syntax Phonology Interface, *Linguistic Inquiry*, to appear.
- Swerts, M., C. Avesani and E. Krahmer (1999), Reaccentuation or deaccentuation: a comparative study of Italian and Dutch. *Proc. 14th ICPHS*, San Francisco, 1541-144.
- Swerts, M., D. Litman and J. Hirschberg (2000), Corrections in spoken dialogue systems, *Proc. ICSLP 2000*, Beijing, China, vol. II, 615-619.
- Swerts, M. and E. Krahmer, in progress. Reconstructing Dialogue History (working title).
- Swerts, M., M. Taniguchi and Y. Katagiri (2000), Prosodic marking of information status in Tokyo Japanese, *Proc. ICSLP 2000*, Beijing, China, vol. I, 78-82.
- Vallduví, E. (1990), *The Information Component*, Ph.D. Dissertation, University of Pennsylvania.
- Weegels, M. (1999), Users' (mis)conceptions of a voice-operated train travel information service. IPO Annual Progress Report, Eindhoven, The Netherlands, pp. 45-52.